



King's Research Portal

DOI:

[10.1016/j.psychres.2013.03.007](https://doi.org/10.1016/j.psychres.2013.03.007)

Document Version

Early version, also known as pre-print

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Brohan, E., Clement, S., Rose, D., Sartorius, N., Slade, M., & Thornicroft, G. (2013). Development and psychometric evaluation of the Discrimination and Stigma Scale (DISC). *Psychiatry Research*, 208(1), 33-40. [10.1016/j.psychres.2013.03.007](https://doi.org/10.1016/j.psychres.2013.03.007)

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Title: Development and psychometric evaluation of the Discrimination and Stigma Scale (DISC)

Abstract

Mental illness is associated with unfair treatment in a number of areas of life. There is currently no psychometrically validated measure that has been developed to specifically focus on such experienced discrimination. This study aimed to finalise the Discrimination and Stigma Scale (DISC) and establish its psychometric properties. DISC was further developed using: 1) service user and interviewer focus groups; 2) reading ease testing; and 3) cognitive debriefing interviews. The revised scale then underwent psychometric testing to establish the following properties: reliability; validity; precision; acceptability; and feasibility. The final 22-item DISC demonstrated good psychometric properties (n=86) including inter-rater reliability (weighted kappa range: 0.62-0.95), internal consistency ($\alpha=0.78$) and test-retest reliability (n=46) (weighted kappa range: 0.56-0.89). Feasibility, validity and acceptability were also established. In conclusion, the 22-item DISC is recommended for use in measuring experienced stigma and discrimination. Additional work to develop a measure of anticipated stigma is recommended.

Key words: Discrimination and stigma scale (DISC); psychometric validation; scale development; stigma; mental illness

1. Introduction

Stigma is defined as a characteristic which individuals possess (or are believed to possess) that conveys an identity which is devalued in a particular social context (1). Mental illness is associated with devaluation in a number of social contexts including: the workplace; healthcare settings; acting as a parent; and personal relationships (2-5). This devaluation can be targeted towards the stigmatized individual or those close to them, in what is termed as courtesy stigma (6). Three elements of personal stigma can be considered: 1) perceived stigma or the belief that the public hold negative attitudes towards people with a mental health problem, 2) experienced stigma or reported instances of unfair treatment or discrimination due to having a mental health problem and 3) self-stigma or adopting a stigmatised view of oneself (7). Stigma research has largely focused on the measurement of perceived stigma. A recent review of studies using survey-based measures (n=52), reported that 79% used a measure of perceived stigma, 46% a measure of experienced stigma and 33% a measure of self-stigma (8). This review further suggests that, although several survey measures have addressed aspects of experienced stigma, there is currently no psychometrically validated measure that was developed with a specific focus on understanding the scope and content of these experiences.

The Discrimination and Stigma scale (DISC) was developed to address this gap. It is based on the definition of Thornicroft et al, 2007 who present stigma as an overarching term including three elements: 1) problems of knowledge (ignorance or misinformation); 2) problems of attitudes (prejudice); and 3) problems of behaviour (discrimination) (9). It focuses on the third aspect of stigma: problems of behaviour or discrimination. It is an interview-based measure which collects qualitative and quantitative data to provide a rating of the degree to which discrimination has been experienced in various areas of life including work, relationships, parenting, housing, leisure, and religious activities.

A preliminary version of DISC was developed as part of the International Study on Discrimination and Stigma Outcomes (INDIGO) (10). As part of the scale development, face and content validity were established through a literature review, Delphi consultation and pilot testing of the draft scale within research teams at 28 participant study sites, in 27 countries. Twenty-five interviews were conducted at each site (total n=732) with five of the interviews at each site audio-taped, transcribed verbatim, translated into English and qualitatively analysed by members of the study team (11). The qualitative and quantitative analysis of the data collected suggested that the scale may benefit from further development work.

1.1. Aims

This study aims to:

- 1) Complete the developmental work to maximise the acceptability and feasibility of DISC, leading to a finalised version of the scale (Phase 1)
- 2) Establish the psychometric properties of the revised DISC (Phase 2)

2. Methods

2.1. Methods for completing the development of DISC

Firstly, the corrections indicated by the INDIGO data analyses were incorporated, as described elsewhere (10, 11). This was followed by three stages of evidence gathering to support further development: 1) service user and interviewer focus groups (n=4 groups); 2) reading ease testing; and 3) cognitive debriefing. Evidence from stage 1 was used to create a draft DISC, which was considered in stages 2-3.

2.1.1. Stage 1: Service user and interviewer focus groups

Two semi-structured focus groups were conducted with mental health service users. Eligible participants were identified by the clinical team at a day centre service. Participants were asked to complete DISC at the beginning of the focus group and then discuss aspects of the scale including: overall ease of completion; relevance of items and response options, time taken to complete and recommendations for improvement. Two focus groups were also separately conducted with interviewers who had experience of using DISC as part of other studies run by colleagues at the Institute of Psychiatry. Group discussion focused on experience of using the scale and recommendations for improvement.

2.1.2. Stage 2: Reading ease testing

The Flesch Reading Ease score and Flesch-Kincaid Grade level were assessed using MS Word. These are widely used tools which assess readability based on the syllabic and sentence structure of the text (12). The Flesch Reading Ease score ranges from 0–100 with higher scores being easier to read. The Flesch-Kincaid Grade level provides an indication of the US educational grade to which the material is most appropriate (range 0 to 17)(13).

2.1.3. Stage 3: Cognitive debriefing

Cognitive debriefing involves a small interview study, providing qualitative data on the mental processes that respondents use to answer questions (14). Once changes had been made following the recommendations of stages 1 and 2, cognitive debriefing interviews were conducted as a check to ensure conceptual clarity and ease of meaning in the final DISC. Five individuals were recruited from a day centre service using the methods described in the service user focus groups.

2.2. Methods for the psychometric evaluation of DISC

2.2.1. Design

A cross sectional study design was used with participants interviewed at one point in time. A sub-sample of participants also completed the DISC again 7-14 days following initial administration to establish the test-retest reliability. This study, and the earlier work described in stages 1-3 above, received National Research Ethics Service (NRES) approval from the Camden and Islington Community Local Research Ethics Committee (REC ref: 08/H0722/40).

2.2.2. Sample

A sample size of 90 was chosen as it is sufficient to establish that the inter-rater reliability is at least 0.7 (assuming that the true level is 0.8) (15). Sample size was calculated based on inter-rater reliability as this was the property which required the largest sample. The sample of 90 was also sufficient to establish the other properties

under consideration. Convenience sampling was used. All participants were aged 18 years or over and had used a community mental health team in the past 12 months.

2.2.3. Procedure

Interviews were conducted by a team of six researchers. Each researcher participated in a half day training session lead by EB. During the participant interviews, DISC responses were digitally recorded. A second member of the research team listened to the interview and recorded a score for each DISC item. This was compared with the original interview scores to calculate inter-rater reliability. Every second participant in this study was asked to repeat the DISC interview again 7-14 days following initial administration. The study measures are detailed below:

2.2.4. Measures

Discrimination and Stigma Scale (DISC)

The revised DISC is a 35 item, interview-based, measure. All items are scored on a four-point Likert Scale ranging from 0=not at all to 3=a lot. It comprises a global scale and four subscales, each of which is scored separately. The four subscales are: 1) Unfair treatment (22 items); 2) Stopping self (4 items); 3) Overcoming stigma (2 items) and 4) Positive treatment (7 items). Both a mean and a total score are calculated for each subscale and the global scale. This allows both the typical level of stigma in each applicable area of life, and its spread over the different areas to be present

2.2.4.1. The Stigma Scale (SS)

The SS is a 28-item self-complete measure. It has three subscales: disclosure (11 items), discrimination (12 items) and positive aspects (5 items) (5). The scale has good test-retest reliability (kappa range 0.49-0.71) and internal consistency ($\alpha=0.87$).

2.2.4.2. Internalised Stigma of Mental Illness Scale (ISMI)

The ISMI is a 29-item self-complete measure that assesses mental health service users' experience of internalised stigma (2). It is composed of five subscales: Alienation, Stereotype Endorsement, Perceived Discrimination, Social Withdrawal and Stigma Resistance. Strong internal consistency ($\alpha=0.90$) and test-retest reliability ($r=0.92$) have been reported.

2.2.4.3. Brief Psychiatric Rating Scale (BPRS)

The BPRS measures psychiatric symptomatology, including positive symptoms, general psychopathology and affective symptoms (16). The 18-item version of the scale was used in this study (17). Items 1-10 are rated by the participant during an interview, while items 11-18 are rated by the researcher following observation of the participant. Each item is rated on a seven point scale anchored at 1=not present and 7=extremely severe.

2.2.4.4. Global Assessment of Functioning (GAF)

The GAF comprises two numeric scales (18). Each is rated by an observer from 0 (most severe) to 100 (least severe). One (GAF-D) asks the rater to consider the level of disability, and the other (GAF-S) symptoms. The GAF has been widely used and is a reliable and valid measure of functioning (19).

2.3. Data analysis

2.3.1 Completing the development of DISC

Verbatim transcripts from the four focus groups were analysed for recurrent themes. Suggestions of changes were grouped under these themes and discussed within the study team. The reading ease scores and cognitive debriefing results were also considered. This work focused on maximising the acceptability and feasibility of DISC. Acceptability describes the extent to which a scale is targeted towards the intended population. The suitability of the wording of the survey to the target audience was established using the service user focus groups, service user research report, reading ease testing and cognitive debriefing interviews. Acceptability can also be used to describe the quality of data as assessed by the completeness of data and score distributions (20). This property will be established as part of the psychometric evaluation of DISC in the next section. Feasibility asks whether the scale is easy to administer and process (21). This was considered by conducting the interviewer focus groups, where individuals with experience of using DISCv11 were asked to discuss aspects related to feasibility. Further aspects of feasibility are also considered as part of the psychometric evaluation in the next section.

2.3.2. Psychometric evaluation of DISC

Analysis was performed using SPSS version 15 (22) and Stata version 9.2 (23). To determine whether the variables to be used in the analysis were sufficiently normally distributed, histograms were examined for all continuous variables. Skew and kurtosis values were examined and a Kolmogorov-Smirnov test was performed. Outliers were identified using the z-score criterion of ± 3.29 , with those cases violating this threshold excluded from analysis. Descriptive statistics for the socio-demographic and illness-related variables were calculated.

2.3.2.1. Scoring

Descriptive statistics were calculated for all DISC scale and subscale scores. Positive treatment is reverse coded so a high score on this variable indicates a lack of positive treatment or occasions when the person received help or support not available to others, as a result of their mental health problem. A strategy for the interpretation of scores was applied using the DISC midpoint of 1.5. In this interpretation four categories were used to understand DISC mean scores: <1 minimal discrimination; 1-1.5 low discrimination; 1.5-2 moderate discrimination; and 2+ high discrimination. This scoring strategy was based on a previous approach in studies using the ISMI, which represented a high level of self-stigma as an average score above the midpoint of 2.5 (24, 25).

2.3.2.2. Reliability

The reliability of DISC was assessed by considering: 1) consistency over sub-components (internal consistency); 2) consistency over raters (inter-rater reliability) and 3) consistency over time (test-retest reliability). Internal consistency was assessed using Cronbach's α with a criterion of $\alpha \geq 0.70$ indicative of appropriate internal consistency for each subscale (26). $\alpha > 0.90$ were also flagged, as this may indicate item redundancy. Lin's concordance statistic (ρ_c) was used to calculate the overall inter-rater reliability and test-retest reliability for each subscale and the total DISC score (27). This was performed using the 'concord' command in Stata. Key demographic variables were examined to test for differences between those who

were included in the retest study and those not included in the retest study. Due to the normal distribution, the t-test was conducted in the case of continuous variables (age, social contact) and the chi-square test or Fisher's exact test was used in the case of categorical data (all other variables). A criterion of Lin's $\rho_c \geq 0.70$ was used to indicate acceptable reliability for all scores. A weighted kappa score was also calculated for each item pair to provide details on the test-retest and inter-rater reliability of individual items. A criterion of weighted kappa ≥ 0.4 was used to indicate acceptable item level reliability.

2.3.2.3. Validity

The following aspects of validity were established in this study: 1) within-scale validity; and 2) convergent and discriminant validity. Within-scale analysis of validity provides evidence that a single construct is being measured and that items can be combined to form a total score. This is assessed based on internal consistency, as established in the reliability analysis. The correlations between scales were also examined to assess the degree to which they measure related aspects of the construct. Moderate correlations were taken as the criterion for this (20). Convergent and discriminant validity are specific forms of construct validity where it is hypothesised that the scale under consideration will have a stronger association with certain variables and a weaker association with other variables. The convergent validity of DISC subscales was examined as specified below. A significant moderate to strong correlation (0.3 or greater) was taken as the criterion for convergent validity. Terwee and colleagues propose that construct validity is established if at least 75% of results are in accordance with the hypothesised relationships (28). For DISC the following relationships are hypothesised:

1. DISC Unfair Treatment subscale will have a significant association with the Discrimination subscale of the SS and the Discrimination Experiences subscale of the ISMI
2. DISC Stopping Self subscale will have a significant association with the Disclosure subscale of the SS and the Social Withdrawal subscale of the ISMI
3. DISC Overcome Stigma subscale will have a significant association with the Positive Aspects subscales of the SS and the Stigma Resistance subscale of the ISMI
4. DISC Positive Discrimination subscale will have a significant association with the Positive Aspects subscales of the SS and the Stigma Resistance subscale of the ISMI
5. DISC Total Score will have a significant association with the SS Total score and SS Discrimination score.

Divergent validity was measured by assessing the relationship between DISC subscales and Total score with the demographic variable of gender. There is currently a lack of evidence on the relationship between gender and discrimination of mental health problems (29). This variable was selected as a measure of divergent validity as it is expected that DISC scores will not have a significant association with gender, as established in the previous INDIGO study (10). No significant difference in DISC scores by gender will be taken as the appropriate criterion.

2.3.2.4. Precision

This considers the appropriateness of the scaling assumptions i.e. how well each item fits within its proposed scale. Corrected item-total correlations < 0.30 were used

to indicate unacceptable fit of the items with the scale total score (28). Each item was also correlated with its own scale total and with other subscale totals. To maintain the precision of the subscales, each item needed to correlate more highly with its own subscale than with other DISC subscales.

2.3.2.5. Acceptability

Acceptability describes the quality of data as assessed by the completeness of data and score distributions (20). The following aspects of acceptability were established: 1) Maximum endorsement frequencies (MEF); and 2) Aggregate adjacent endorsement frequencies (AEF). To consider MEF, the n(%) of respondents who endorse each response category for each item was presented. MEF >80% in a particular category indicates that the item may need further consideration. AEF criterion is violated when two or more adjacent scale points on an item show <10% of the responses (30).

2.3.2.6. Feasibility

Feasibility was considered by noting the time taken to conduct the DISC interview. A time of greater than 30 minutes was considered indicative of an overly long interview-based measure.

3. Results

3.1. Results of completing the development of DISC

Thirteen people participated in the service user focus groups and twelve in the interviewer focus groups. 39% of participants were male. The mean age of participant was 46.8 (SD=13.42). 80% reported some personal experience of stigma. Participants in the service user focus groups most commonly reported the following diagnoses: depression (n=3); schizophrenia (n=4); and bipolar disorder (n=6). In the interviewer focus groups participants had conducted an average of 5.67 DISC interviews (SD=4.62). The recurrent themes identified by analysing the focus group transcripts suggested five areas of change: 1) reduction of complexity; 2) change phrase 'treated differently from other people'; 3) change phrase 'your diagnosis of mental illness'; 4) change response options; and 5) wording of individual questions. After these changes were incorporated, the Flesch-Kincaid Grade level was reduced from 13.2 to 7.4 and Flesch Reading Ease score increased from 36.2 to 65.7, suggesting that the modifications were successful in reducing the complexity of the scale. The cognitive debriefing interviews further support acceptability of the revised DISC. All items were well understood by the 5 participants and no further areas of clarification were identified. The finalised DISC is a 35-item scale as described in the methods for psychometric evaluation of DISC.

3.2. Results of psychometric evaluation of DISC

3.2.1. Demographic characteristics

86 people took part in the study. Their socio-demographic and clinical characteristics are reported in Table 1. below. The BPRS mean score indicates that participants can be classified as moderately ill (31). 39.8% of those included in this analysis (n=83) fall into this category with a score between 41 and 52. 24.1% were categorised as mildly ill with scores of 31-40, 19.3% were categorised as markedly ill with scores of 53 or greater, and the remaining 16.9% were classified as borderline ill with scores of 30 or less. This is in keeping with the reported GAF scores which suggest that the

sample have a moderate level of symptoms and moderate difficulty in social or occupational functioning (18).

[Insert Table 1 about here]

Figure 1 shows the frequency of endorsement for each DISC item.

[Insert Figure 1 about here]

Table 2 reports on the stigma measures. For DISC, all sub-scale scores fell in the low stigma category (scores of 1-1.5), with the exception of Unfair Treatment which was in the minimal stigma category and Positive Treatment which was in the high stigma category. On average participants reported experiencing stigma in 16.55 of the 35 aspects of life in DISC. All but one participant endorsed at least one aspect in life in which they had experienced stigma (98.8%). The ISMI scores displayed are in keeping with published scores for a similar population e.g. Lysaker and colleagues (2007), report the following: Alienation 2.31 (SD 0.65); Stereotype endorsement 1.99 (SD 0.54); Discrimination experience 2.42 (SD 0.69); Social withdrawal 2.30 (SD 0.66); and Stigma resistance 2.17 (SD 0.52) (32). The displayed SS scores were also in keeping with published scores, as reported by King and colleagues (2007): total SS 62.6 (SD 15.4); Discrimination 29.1 (SD 9.5); Disclosure 24.7 (SD 8.0); and Positive aspects 8.8 (SD 2.8) (5).

[Insert Table 2 about here]

The degree to which the criterion for each psychometric property was established is summarised in Table 3. below. Each property will now be discussed.

[Insert Table 3 about here]

3.2.2. Reliability

3.2.2.1. Internal consistency

Cronbach's α for DISC was 0.78, and removal of any item did not increase the α value to greater than 0.79. Sub-scale analysis showed Cronbach's α for Unfair Treatment was 0.82, for Stopping Self was 0.66 and for Positive Treatment was 0.67. Values were not increased by the deletion of any items.

3.2.2.2. Inter-rater reliability

Lin's $\rho_c=0.89$, $p<0.001$. For individual items the weighted kappa statistic ranged from 0.62 (dating or intimate relationships) to 0.97(stopped self applying for education). The overall inter-rater reliability for the Unfair Treatment subscale Lin's $\rho_c=0.87$, $p<0.001$, for the Stopping Self subscale Lin's $\rho_c=0.91$, $p<0.001$, for Overcoming stigma Lin's $\rho_c=0.78$, $p<0.001$ and for Positive Treatment Lin's $\rho_c=0.91$, $p<0.001$.

3.2.2.3. Test-retest reliability

There were no significant differences between those who took part in the retest study and those who did not, on key demographic including gender, age, employment status, ethnicity and diagnosis. DISC (n=44) had an overall test-retest reliability Lin's

$\rho_c=0.88$, $p<0.001$. For individual items the weighted kappa statistic ranged from 0.45 (treated more positively by mental health staff) to 0.89 (treated unfairly in keeping a job). The overall test-retest reliability for the unfair treatment subscale ($n=46$) Lin's $\rho_c=0.89$, $p<0.001$, for the Stopping Self subscale ($n=45$) Lin's $\rho_c=0.71$, $p<0.001$, for Overcoming Stigma ($n=45$) Lin's $\rho_c=0.56$, $p<0.001$ and for Positive Treatment ($n=45$) Lin's $\rho_c=0.50$, $p<0.001$.

3.2.3. Validity

3.2.3.1. Inter-correlation of DISC subscales

The DISC Total, Unfair Treatment and Stopping Self subscales reached the threshold moderate level (0.3), while the Overcoming Stigma (correlation with Stopping Self $r=0.20$, NS) and Positive Treatment subscales (correlation with Unfair Treatment $r=0.14$, NS, correlation with Stopping Self $r=-0.02$, NS) did not meet this criterion.

3.2.3.2. Convergent Validity

DISC Unfair Treatment subscale showed adequate convergent validity with the Discrimination subscale of the SS ($r=0.54$, $p<0.001$) and the Discrimination Experience subscale of the ISMI ($r=0.31$, $p<0.001$). DISC Stopped Self subscale was significantly correlated with the Disclosure subscale of the SS ($r=0.25$, $p<0.05$) and the Social Withdrawal subscale of the ISMI ($r=0.23$, $p<0.05$). Neither of these correlations reached the threshold of 0.3. DISC Overcome Stigma subscale was significantly correlated with the Positive Aspects subscale of the SS ($r=-0.24$, $p<0.05$) and the Stigma Resistance subscale of the ISMI ($r=0.29$, $p<0.001$). Neither met the threshold of 0.3. DISC Positive Treatment subscale was not significantly correlated with the Positive Aspects subscales of the SS ($r=-.20$, NS) or the Stigma Resistance subscale of the ISMI ($r=-0.06$, NS). The DISC Total score was significantly correlated with the SS Total score ($r=0.26$, $p<0.05$) and Discrimination score ($r=0.38$, $p<0.001$), with the later reaching threshold.

3.2.3.3. Divergent Validity

Divergent validity was assessed by considering the association between gender and DISC Total score. No significant relationship was found (mean score males=15.53 (SD=6.55), mean score females=17.97(SD=4.68, $t=-1.89$, $p=0.063$).

3.2.4. Precision

3.2.4.1. Corrected item-total correlations

The total corrected item-total correlation (r_s) for the Unfair Treatment subscale was .38. Correlations between items and the subscale corrected item totals ranged from $r_s=.55$ for 'treated unfairly making or keeping friends' to $r_s=.24$ for 'treated unfairly in marriage or divorce', which had a correlation lower than the predefined threshold of $r_s=.3$. The total corrected item-total correlation for the Stopping Self subscale was $r_s=.66$. Correlations between items and the subscale total ranged from $r_s=.50$ for 'stopped yourself from having a close personal relationship' to $r_s=.66$ for 'stopped yourself from applying for education or training courses'. No items had a correlation lower than the predefined threshold of $r_s=.3$. The total corrected item-total correlation for the Overcoming Stigma subscale was $r_s=.20$. Correlations between the item and subscale total was $r_s=.20$ for each item. No items met the predefined threshold. The

total corrected item-total correlation for the positive treatment subscale was $r_s=.67$. Correlations between items and the subscale total ranged from $r_s=.59$ for 'treated more positively by mental health staff' to $r_s=.68$ for 'treated more positively by your family'. No items had a correlation lower than the predefined threshold.

3.2.4.2. Correlation with own subscale and others

All items, with the exception of 'use your personal skills or abilities in coping with stigma', correlated more highly with their own subscale than any of the other three subscales. For the Unfair Treatment subscale, correlations of items with the subscale ranged from $r=0.25$, $p<0.05$ for 'treated unfairly in marriage or divorce' to $r=0.59$, $p<0.01$ for 'treated unfairly making or keeping friends'. For the Stopping Self subscale, all items correlated more highly with this subscale rather than any of the others. Correlations of items with the subscale ranged from $r=0.36$, $p<0.01$ for 'stopped yourself from applying for education or training courses' to $r=0.57$, $p<0.01$ for 'stopped yourself from having a close personal relationship'. For the Overcoming Stigma subscale, 'made friends with people who don't use mental health services was correlated' with the subscale total at $r=0.33$, $p<0.01$ while 'use your personal skills or abilities in coping with stigma' was correlated at $r=0.31$, $p<0.01$. For the Positive Treatment subscale, correlations of items with the subscale ranged from $r=0.29$, $p<0.01$ for 'treated more positively in employment' to $r=0.56$, $p<0.01$ for 'treated more positively by mental health staff'.

3.2.4.3. Acceptability

Several of the items had a high endorsement of the non applicable response e.g. 'treated unfairly in your role as a parent to your children' at 67.1% and 'treated unfairly in marriage or divorce' at 58.8%. The MEF criterion was not violated as no response categories contained $\geq 80\%$ of responses. There were no items which violated the AEF when considering the adjacent categories of 'not at all' and 'a little'. There were three items which violated the AEF when considering the adjacent categories of 'a little' and 'moderately'; 'treated unfairly in your role as a parent to your children'; 'treated unfairly in marriage or divorce'; 'treated unfairly in starting a family or having children'. Of these items, the non applicable response option was frequently endorsed at 67.1%, 58.8% and 42.9% respectively. One item violated the AEF when considering the adjacent categories of 'moderately' and 'a lot'; 'treated unfairly in religious practices'. This also had a large endorsement of the 'not at all' category at 60.7% of responses.

3.2.5. Feasibility

DISC completion time ranged from 5.0 minutes to 85.3 minutes. When two outlying cases (82.9 minutes and 85.3 minutes) were removed the mean completion time was 29.1 minutes ($SD=12.8$). The median completion time was 28.4 minutes.

4. Discussion

This study conducted further developmental work and evaluated the psychometric properties of DISC. The reliability of DISC was satisfactorily established, including internal consistency, test-retest reliability and inter-rater reliability. The criterion for feasibility was also established. Elements of validity, precision, and acceptability were established, and areas leading to the further improvement of these properties were also highlighted. The Unfair Treatment subscale appeared robust, fully meeting the validity and precision criteria. The only area in which it did not meet the threshold

was the AEF in which four items violated the criterion. This violation is not necessarily a cause for concern however, as these items detail areas of life which were not applicable to all participants including parenting, marriage and religion. The high proportion of non applicable responses for these items was a contributing factor to the violation of the AEF. These items were considered sufficiently important to include despite their lack of universal applicability. Therefore there is a conceptual justification for their inclusion despite this psychometric violation. This strong psychometric profile leads to a recommendation for the 22-item subscale to be used as a stand alone scale to measure experienced stigma.

The Overcoming Stigma and Positive Treatment subscales did not correlate moderately with the Stopping Self and Unfair Treatment subscales, suggesting that they may be measuring a different construct to that assessed in the other subscales. The lack of convergent validity between the Stopping Self subscale and the Disclosure subscale of SS and Social Withdrawal subscale of ISMI suggests that further work is needed on this subscale if it is to represent these constructs adequately. Further work on these three subscales is recommended to improve their validity and precision. However, they are sufficiently reliable, acceptable and feasible for current use. Further work is underway to develop a new measure of anticipated discrimination – the Questionnaire on Anticipated Discrimination (QUAD). This measure will be designed with the aim of using it to replace the Stopping Self, Overcoming Stigma and Positive Subscales of DISC. These subscales will be used as the initial basis of the new scale. Therefore it is recommended to drop these elements from DISC, and use it as a 22-item measure of experienced stigma, as currently represented by the Unfair Treatment subscale. Further work will also be necessary to evaluate the psychometric properties of DISC in additional clinical populations or in cultural groups other than those reported in this paper.

4.1. Strengths and limitations

The psychometric evaluation of DISC is strengthened by building on the assessment of elements of acceptability and feasibility, as established in the further development work. The 22-item scale contributes evidence to support the evaluation of outcomes as part of social inclusion interventions or anti-stigma campaigns, as indicated by the Medical Research Council's guidance on developing and evaluating complex interventions (33).

This study is limited as data collection was not powered to perform factor analysis. This decision was taken as this psychometric requirement can be fulfilled by performing secondary analysis on an existing dataset. This additional dataset was however not sufficient to conduct the analyses described here as no additional measures were administered, multiple ratings were not performed and data collection was at one time point only. Secondary analysis of this dataset will allow the factor structure of the DISC to be established, complementing the analyses performed in this paper. Secondary analysis of the additional dataset will also allow an item response theory approach to be taken, particularly the investigation of Differential Item Functioning. ANOVA will be conducted for each item to examine uniform and non-uniform Differential Item Functioning across each level of the 'person factor' (employment status) and across different levels of the DISC constructs (class intervals). Again, the sample size of this current study is not sufficient to perform these analyses.

4.2. Conclusions

In conclusion, the 22-item DISC is a reliable, valid, precise, acceptable, and feasible measure for use in assessing experienced discrimination. The use of this scale is recommended as an evaluation tool. The Stopping Self, Overcoming Stigma and Positive Treatment subscales require further modification before they can be recommended as meeting all psychometric requirements therefore they are not included in the recommended version of DISC. Further work is underway to develop these subscales into a stand alone measure.

Acknowledgments

Data collection assistance was given by Amy MacDonald, Gaby Illingsworth, Zoe Given-Wilson, Eleanor Parker, Manuela Jarrett, Tanya Graham, Georgia Black and Eleanor Lewis-Holmes. Dr Morven Leese provided statistical advice. This paper presents independent research commissioned by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research scheme (RP-PG-0606-1053). The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. GT is also funded through a NIHR Specialist Mental Health Biomedical Research Centre at the Institute of Psychiatry, King's College London and the South London and Maudsley NHS Foundation Trust.

1. Crocker J, Major B, Steele C. Social stigma. In: Gilbert DT, Fiske ST, Lindzey G, editors. *Handbook of Social Psychology*. 2. 4th ed. Boston: McGraw-Hill; 1998. p. 504-33.
2. Ritsher JB, Otilingam PG, Grajales M. Internalized stigma of mental illness: Psychometric properties of a new measure. *Psychiatry Res*. 2003 2003;121(1):31-49.
3. Wahl OF. Mental Health consumers experience of stigma. *Schizophr Bull*. 1999 1999;25(3):467-78.
4. Link BG, Struening EL, Rahav M, Phelan JC, Nuttbrock L. On stigma and its consequences: evidence from a longitudinal study of men with dual diagnosis of mental illness and substance abuse. *J Health Soc Behav*. 1997 1997;38:177-90.
5. King M, Dinos S, Shaw J, Watson R, Stevens S, Passetti F, et al. The stigma scale: development of a standardised measure of the stigma of mental illness. *Br J Psychiatry*. 2007 2007;190:248-54.
6. Goffman E. *Stigma: Notes on the Management of Spoiled Identity*. Harmondsworth, Middlesex: Penguin Books; 1963.
7. Yanos PT, Roe D, Markus K, Lysaker PH. Pathways between internalised stigma and outcomes related to recovery in schizophrenia spectrum disorders. *Psychiatr Serv*. 2008 2008;59(12):1437-42.
8. Brohan E, Slade M, Clement S, Thornicroft G. Experiences of mental illness stigma, prejudice and discrimination: a review of measures. *BMC Health Serv Res*. 2010 2010;10(1):80.
9. Thornicroft G, Rose D, Kassam A, Sartorius N. Stigma: ignorance, prejudice or discrimination? *Br J Psychiatry*. 2007 3/2007;190:192-3.
10. Thornicroft G, Brohan E, Rose D, Sartorius N, Leese M, Group TIS. Global pattern of experienced and anticipated discrimination against people with schizophrenia: a cross-sectional survey. *The Lancet*. 2009 2009;373(9661):408-15.
11. Rose D, Willis R, Brohan E, Sartorius N, Thornicroft G, Villares C, et al. Reported Stigma and Discrimination by People with a Diagnosis of Schizophrenia. *Epidemiology and psychiatric sciences*. 2011 2011;20:193-204.
12. Kalk NJ, Pothier DD. Patient information on schizophrenia on the internet. *The Psychiatrist*. 2008 11/1/2008;32(11):409-11.
13. Flesch R. *The Art of Readable Writing*. 2nd ed: Harper & Row; 1974 1974.
14. Ojanen V, Gogates G. A briefing on cognitive debriefing. *Good Clinical Practice Journal*. 2006 2006;13(12):25-9.
15. Walter SD, Eliasziw M, Douglas JA. Sample size and optimal designs for reliability studies. *Stat Med*. 1998 1998;17:101-10.

16. Overall JE, Gorham DR. The Brief Psychiatric Rating Scale. *Psychol Rep.* 1962 1962;10:799-812.
17. Lukoff D, Neuchterlein KH, Ventura J. Appendix A. Manual for expanded brief psychiatric rating scale (BPRS). *Schizophr Bull.* 1986 1986;12:594-602.
18. Endicott J, Spitzer R, Fleiss J, Cohen J. The Global Assessment Scale. *Arch Gen Psychiatry.* 1976 1976;33:765-71.
19. Jones SH, Thornicroft G, Coffey M, Dunn G. A brief mental health outcome scale-reliability and validity of the Global Assessment of Functioning (GAF). *The British Journal of Psychiatry.* 1995 5/1/1995;166(5):654-9.
20. Lamping DL, Schroter S, Marquis P, Marrel A, Duprat-Lomon I, Sagnier PP. The Community-Acquired Pneumonia Symptom Questionnaire* : A New, Patient-Based Outcome Measure To Evaluate Symptoms in Patients With Community-Acquired Pneumonia. *Chest.* 2002 9/1/2002;122(3):920-9.
21. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess.* 1998 1998;2(14).
22. Inc. S. SPSS for Windows, Rel 15.0. Chicago: SPSS Inc; 2006.
23. StataCorp. Stata Statistical Software: Release 9.2. College Station, TX: StataCorp LP; 2005.
24. Brohan E, Elgie R, Sartorius N, Thornicroft G. Self-stigma, empowerment and perceived discrimination among people with schizophrenia in 14 European countries: The GAMIAN-Europe study. *Schizophr Res.* 2010 2010;122(1):232-8.
25. Brohan E, Gauci D, Sartorius N, Thornicroft G. Self-stigma, empowerment and perceived discrimination among people with bipolar disorder or depression in 13 European countries: The GAMIAN-Europe study. *J Affect Disord.* 2011 3/2011;129(1-3):56-63.
26. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika.* 1951 1951;16:297-334.
27. Lin LIK. A concordance correlation coefficient to evaluate reproducibility. *Biometrics.* 1989 1989;45:255-68.
28. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007 1/2007;60(1):34-42.
29. Angermeyer MC, Dietrich S. Public beliefs about and attitudes towards people with mental illness: a review of population studies. *Acta Psychiatr Scand.* 2006 2006;113(3):163-79.
30. The Whoqol Group. The World Health Organization quality of life assessment (WHOQOL): Development and general psychometric properties. *Soc Sci Med.* 1998 6/15/1998;46(12):1569-85.

31. Leucht S, Kane JM, Kissling W, Hamann J, Etschel E, Engel R. Clinical implications of Brief Psychiatric Rating Scale scores. *The British Journal of Psychiatry*. 2005 10/1/2005;187(4):366-71.
32. Lysaker PH, Roe D, Yanos PT. Toward understanding the insight paradox: Internalized stigma moderates the association between insight and social functioning, hope, and self-esteem among people with schizophrenia spectrum disorders. [References]. *Schizophr Bull*. 2007 2007;33(1):192-9.
33. Craig P, Dieppe P, MacIntyre S, Mitchie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. *Br Med J*. 2008 2008;337:979-83.